

educ_lin.R

datalab

2023-06-16

```
#####  
##### Linear Polynomial #####  
##### Robustness on Education  
##### December 15, 2022
```

```
rm(list=ls())  
library(foreign)  
library(plyr)  
library(readstata13)  
library(multiwayvcov)  
library(sandwich)  
library(lmtest)  
library(stargazer)
```

```
data=read.csv("~/Dropbox/Personal Research 2017/replications/karn_nov16.csv")  
names(data)
```

```
## [1] "X.1" "dist_name" "vilname91" "v1" "dist_code"  
## [15] "phc_cntr" "hc_cntr" "fpc_cntr" "tb_cntr" "nh_cntr"  
## [29] "st_town" "agri_land" "near_town" "circl_code" "m_pop"  
## [43] "m_sc" "f_sc" "tot_st" "m_st" "f_st"  
## [57] "ngmf_char" "ngmfprwmedd" "ngmfprwothd" "ngmfprwnod" "ngmftrpr"  
## [71] "taptr" "tapuntr" "hp" "covwell" "uncovwell"  
## [85] "phs_cntr" "prhsc" "stname" "stname1991" "d_name"  
## [99] "all_hosp" "area_na_cu" "ayu_disp" "ayu_hosp" "canal_govt"  
## [113] "ind_sch" "lake" "m_home" "m_sch" "nw_fac"  
## [127] "power_oth" "power_supl" "p_sch" "p_t_fac" "rang_mcw"  
## [141] "s_sch" "s_s_sch" "tot_exp" "tot_inc" "tr_sch"  
## [155] "gov_ps_n" "pr_ps_n" "gov_ms_n" "pr_ms_n" "gov_secs_n"  
## [169] "nviltmsna" "nviltsecs" "nviltsecsna" "power" "hplost"  
## [183] "pnt_fac" "power_ea" "power_eag" "power_edea" "power_eo"  
## [197] "medfac" "rangmed" "tot_hh" "pucca_binary" "kucha_binary"  
## [211] "dist_fr_town" "tbcl" "tank" "tap" "X"  
## [225] "M_POP" "F_POP" "TOT_L6" "M_L6" "F_L6"  
## [239] "TOT_ILLT" "M_ILLT" "F_ILLT" "TOT_W" "M_W"  
## [253] "F_AGLB" "TOT_MFHH" "M_MFHH" "F_MFHH" "TOT_OTH_W"  
## [267] "M_MRG_AGLB" "F_MRG_AGLB" "T_MRG_HH" "M_MRG_HH" "F_MRG_HH"  
## [281] "NEAR_DIST_border1" "NEAR_ANGLE" "temp_av" "wc2010mt_1" "TerrainRug"
```

```
summary(data$Latitude)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## 13.49 14.28 14.96 15.25 16.21 17.75 138
```

```

summary(data$Longitude)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      74.12  75.26  75.89   75.90  76.48   77.67   138

summary(data$border1)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.000  0.000  1.000   0.599  1.000   1.000   5146

summary(data$border2)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.000  0.000  1.000   0.569  1.000   1.000   6425

summary(data$Slope)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.00  89.98  89.99   87.69  89.99   90.00   138

#####
#### Distances ####

# 20 km
#Distance to Mysore-Bombay Border
rd10.mb=data[which(data$NEAR_DIST_border1<10000),] #20 km

table(rd10.mb$border1)

##
##      0      1
## 559 599

#Distance to Hyderabad-Bombay Border
rd10.hb=data[which(data$NEAR_DIST_border2<10000),] #20 km

table(rd10.hb$border2)

##
##      0      1
## 447 493

#baseline bandwidth
#Mysore
#outcome-high schools
educ.mys10=lm(high_binary~border1+TOT_POP+
              TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.mb) #OLS estimation
summary(educ.mys10)

##
## Call:
## lm(formula = high_binary ~ border1 + TOT_POP + TOT_SC + TOT_ST +
##      Slope + TerrainRug + Latitude + Longitude, data = rd10.mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50730 -0.23118 -0.14076 -0.02466  1.01400
##
## Coefficients:

```

```

##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.310e+00  3.143e+00  -2.326 0.020198 *
## border1     6.471e-03  2.586e-02   0.250 0.802482
## TOT_POP     4.241e-05  1.465e-05   2.894 0.003881 **
## TOT_SC      1.647e-04  4.313e-05   3.818 0.000142 ***
## TOT_ST      1.832e-04  5.413e-05   3.384 0.000741 ***
## Slope       1.927e-03  1.049e-03   1.836 0.066642 .
## TerrainRug  -1.713e-02  6.454e-03  -2.654 0.008059 **
## Latitude    -1.042e-01  7.634e-02  -1.364 0.172715
## Longitude   1.163e-01  4.347e-02   2.676 0.007570 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3817 on 1091 degrees of freedom
## (58 observations deleted due to missingness)
## Multiple R-squared:  0.08441, Adjusted R-squared:  0.0777
## F-statistic: 12.57 on 8 and 1091 DF, p-value: < 2.2e-16
educ.mys10.cl=cluster.vcov(educ.mys10, rd10.mb$dist_name)
educ.mys10.se=sqrt(diag(educ.mys10.cl)) #cluster standard errors

#Hyderabad
#outcome-high schools
educ.hyd10=lm(high_binary~border2+TOT_POP+
              TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd10.hb) #ols estimation
summary(educ.hyd10)

##
## Call:
## lm(formula = high_binary ~ border2 + TOT_POP + TOT_SC + TOT_ST +
##     Slope + TerrainRug + Latitude + Longitude, data = rd10.hb)
##
## Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5951 -0.3104 -0.2214  0.5520  0.9250
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.068e+01  7.146e+00   2.895 0.003890 **
## border2      8.672e-02  3.280e-02   2.644 0.008333 **
## TOT_POP      5.511e-05  1.686e-05   3.268 0.001126 **
## TOT_SC       1.641e-04  4.760e-05   3.447 0.000593 ***
## TOT_ST       6.015e-05  6.268e-05   0.960 0.337456
## Slope        4.957e-04  1.141e-03   0.435 0.663958
## TerrainRug   -8.611e-03  1.808e-02  -0.476 0.634036
## Latitude     1.089e-02  3.911e-02   0.279 0.780689
## Longitude    -2.732e-01  9.930e-02  -2.751 0.006066 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4457 on 892 degrees of freedom
## (39 observations deleted due to missingness)
## Multiple R-squared:  0.05254, Adjusted R-squared:  0.04404
## F-statistic: 6.183 on 8 and 892 DF, p-value: 8.657e-08

```

```

educ.hyd10.cl=cluster.vcov(educ.hyd10, rd10.hb$dist_name)
educ.hyd10.se=sqrt(diag(educ.hyd10.cl)) #cluster standard errors

#10 km
#Distance to Mysore-Bombay Border
rd5.mb=data[which(data$NEAR_DIST_border1<5000),]

table(rd5.mb$border1)

##
## 0 1
## 301 325

#Distance to Hyderabad-Bombay Border
rd5.hb=data[which(data$NEAR_DIST_border2<5000),]

table(rd5.hb$border2)

##
## 0 1
## 260 318

#Mysore
#outcome-high schools
educ.mys5=lm(high_binary~border1+TOT_POP+
             TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd5.mb) #OLS estimation
summary(educ.mys5)

##
## Call:
## lm(formula = high_binary ~ border1 + TOT_POP + TOT_SC + TOT_ST +
##     Slope + TerrainRug + Latitude + Longitude, data = rd5.mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51667 -0.23590 -0.14790 -0.00903  1.03251
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.262e+00  4.294e+00  -1.458  0.14527
## border1      3.615e-02  3.403e-02   1.062  0.28855
## TOT_POP      4.129e-05  2.025e-05   2.039  0.04190 *
## TOT_SC       1.995e-04  6.161e-05   3.239  0.00127 **
## TOT_ST       1.761e-04  7.477e-05   2.355  0.01885 *
## Slope        3.636e-03  1.334e-03   2.725  0.00662 **
## TerrainRug  -1.965e-02  8.827e-03  -2.227  0.02635 *
## Latitude    -6.942e-02  1.145e-01  -0.606  0.54461
## Longitude    9.364e-02  5.824e-02   1.608  0.10841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3876 on 586 degrees of freedom
## (31 observations deleted due to missingness)

```

```

## Multiple R-squared:  0.08691,    Adjusted R-squared:  0.07445
## F-statistic: 6.972 on 8 and 586 DF,  p-value: 8.43e-09

educ.mys5.cl=cluster.vcov(educ.mys5, rd5.mb$dist_name)
educ.mys5.se=sqrt(diag(educ.mys5.cl)) #cluster standard errors

#Hyderabad
#outcome-high schools
educ.hyd5=lm(high_binary~border2+TOT_POP+
             TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd5.hb) #ols estimation
summary(educ.hyd5)

##
## Call:
## lm(formula = high_binary ~ border2 + TOT_POP + TOT_SC + TOT_ST +
##      Slope + TerrainRug + Latitude + Longitude, data = rd5.hb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6099 -0.3314 -0.2181  0.5372  0.9225
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.931e+01  9.673e+00   3.031  0.00256 **
## border2      8.498e-02  4.121e-02   2.062  0.03967 *
## TOT_POP      6.456e-05  2.276e-05   2.836  0.00474 **
## TOT_SC       1.470e-04  6.507e-05   2.259  0.02430 *
## TOT_ST      -3.305e-06  8.321e-05  -0.040  0.96833
## Slope        6.883e-04  1.349e-03   0.510  0.61014
## TerrainRug  -6.382e-03  2.485e-02  -0.257  0.79743
## Latitude     4.141e-02  5.402e-02   0.767  0.44364
## Longitude   -3.929e-01  1.348e-01  -2.915  0.00371 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4534 on 542 degrees of freedom
## (27 observations deleted due to missingness)
## Multiple R-squared:  0.06117,    Adjusted R-squared:  0.04731
## F-statistic: 4.414 on 8 and 542 DF,  p-value: 3.437e-05

educ.hyd5.cl=cluster.vcov(educ.hyd5, rd5.hb$dist_name)
educ.hyd5.se=sqrt(diag(educ.hyd5.cl)) #cluster standard errors

#50 km
#Distance to Mysore-Bombay Border
rd25.mb=data[which(data$NEAR_DIST_border1<25000),]

table(rd25.mb$border1)

##
##      0      1
## 1196 1329

```

```

#Distance to Hyderabad-Bombay Border
rd25.hb=data[which(data$NEAR_DIST_border2<25000),]

table(rd25.hb$border2)

##
## 0 1
## 866 983

#Mysore
#outcome-high schools
educ.mys25=lm(high_binary~border1+TOT_POP+
              TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd25.mb) #OLS estimation
summary(educ.mys25)

##
## Call:
## lm(formula = high_binary ~ border1 + TOT_POP + TOT_SC + TOT_ST +
##     Slope + TerrainRug + Latitude + Longitude, data = rd25.mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5906 -0.2401 -0.1475 -0.0301  0.9862
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.219e+00  1.992e+00  -4.127 3.80e-05 ***
## border1      -7.079e-03  2.164e-02  -0.327  0.7437
## TOT_POP       6.162e-05  1.032e-05   5.968 2.75e-09 ***
## TOT_SC        1.393e-04  2.844e-05   4.897 1.04e-06 ***
## TOT_ST        2.015e-04  3.674e-05   5.485 4.57e-08 ***
## Slope         6.349e-04  6.723e-04   0.944  0.3451
## TerrainRug   -8.238e-03  4.215e-03  -1.954  0.0508 .
## Latitude     -2.555e-02  4.473e-02  -0.571  0.5679
## Longitude     1.144e-01  2.793e-02   4.097 4.33e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3868 on 2395 degrees of freedom
## (121 observations deleted due to missingness)
## Multiple R-squared:  0.08145, Adjusted R-squared:  0.07838
## F-statistic: 26.55 on 8 and 2395 DF, p-value: < 2.2e-16

educ.mys25.cl=cluster.vcov(educ.mys25, rd25.mb$dist_name)
educ.mys25.se=sqrt(diag(educ.mys25.cl)) #cluster standard errors

#Hyderabad
#outcome-high schools
educ.hyd25=lm(high_binary~border2+TOT_POP+
              TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd25.hb) #ols estimation
summary(educ.hyd25)

##

```

```

## Call:
## lm(formula = high_binary ~ border2 + TOT_POP + TOT_SC + TOT_ST +
##      Slope + TerrainRug + Latitude + Longitude, data = rd25.hb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6244 -0.3025 -0.2023  0.5219  0.9732
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.788e+01  4.278e+00   4.179 3.07e-05 ***
## border2      1.117e-01  2.767e-02   4.037 5.64e-05 ***
## TOT_POP      7.207e-05  1.158e-05   6.223 6.06e-10 ***
## TOT_SC       1.666e-04  3.257e-05   5.115 3.48e-07 ***
## TOT_ST       1.235e-04  4.192e-05   2.947 0.00325 **
## Slope        9.867e-04  7.052e-04   1.399 0.16198
## TerrainRug   -1.739e-02  1.166e-02  -1.492 0.13591
## Latitude     4.158e-02  2.478e-02   1.678 0.09352 .
## Longitude   -2.442e-01  5.956e-02  -4.101 4.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.436 on 1778 degrees of freedom
## (62 observations deleted due to missingness)
## Multiple R-squared:  0.06472, Adjusted R-squared:  0.06051
## F-statistic: 15.38 on 8 and 1778 DF, p-value: < 2.2e-16
educ.hyd25.c1=cluster.vcov(educ.hyd25, rd25.hb$dist_name)
educ.hyd25.se=sqrt(diag(educ.hyd25.c1)) #cluster standard errors

#100 km

#Distance to Mysore-Bombay Border
rd50.mb=data[which(data$NEAR_DIST_border1<50000),] #20 km

table(rd50.mb$border1)

##
##      0      1
## 1858 2635

#Distance to Hyderabad-Bombay Border
rd50.hb=data[which(data$NEAR_DIST_border2<50000),] #20 km

table(rd50.hb$border2)

##
##      0      1
## 1540 1905

#baseline bandwidth
#Mysore
#outcome-high schools
educ.mys50=lm(high_binary~border1+TOT_POP+

```

```
TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd50.mb) #OLS estimation
summary(educ.mys50)
```

```
##
## Call:
## lm(formula = high_binary ~ border1 + TOT_POP + TOT_SC + TOT_ST +
##     Slope + TerrainRug + Latitude + Longitude, data = rd50.mb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58971 -0.23116 -0.15145 -0.04807  1.02322
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.100e+00  1.319e+00  -2.351  0.0188 *
## border1      9.091e-03  1.985e-02   0.458  0.6469
## TOT_POP      5.360e-05  7.892e-06   6.792 1.26e-11 ***
## TOT_SC       1.366e-04  2.142e-05   6.379 1.97e-10 ***
## TOT_ST       2.364e-04  2.811e-05   8.409 < 2e-16 ***
## Slope        3.185e-04  4.887e-04   0.652  0.5147
## TerrainRug  -8.881e-03  3.058e-03  -2.904  0.0037 **
## Latitude    -1.503e-03  2.637e-02  -0.057  0.9546
## Longitude    4.194e-02  1.827e-02   2.295  0.0218 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3863 on 4212 degrees of freedom
## (272 observations deleted due to missingness)
## Multiple R-squared:  0.06586,    Adjusted R-squared:  0.06409
## F-statistic: 37.12 on 8 and 4212 DF,  p-value: < 2.2e-16
```

```
educ.mys50.cl=cluster.vcov(educ.mys50, rd50.mb$dist_name)
educ.mys50.se=sqrt(diag(educ.mys50.cl)) #cluster standard errors
```

```
#Hyderabad
```

```
#outcome-high schools
```

```
educ.hyd50=lm(high_binary~border2+TOT_POP+
              TOT_SC+TOT_ST+Slope+TerrainRug+Latitude+Longitude, data=rd50.hb) #ols estimation
summary(educ.hyd50)
```

```
##
## Call:
## lm(formula = high_binary ~ border2 + TOT_POP + TOT_SC + TOT_ST +
##     Slope + TerrainRug + Latitude + Longitude, data = rd50.hb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6870 -0.3137 -0.1990  0.4864  1.0206
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.531e+01  2.481e+00   6.169 7.70e-10 ***
```

```
## border2      1.277e-01  2.446e-02  5.219 1.91e-07 ***
## TOT_POP     9.066e-05  8.535e-06 10.622 < 2e-16 ***
## TOT_SC      1.872e-04  2.337e-05  8.008 1.60e-15 ***
## TOT_ST      1.096e-04  3.009e-05  3.641 0.000275 ***
## Slope       6.615e-04  5.226e-04  1.266 0.205612
## TerrainRug  -9.560e-03  7.632e-03 -1.253 0.210421
## Latitude    7.399e-02  1.492e-02  4.960 7.41e-07 ***
## Longitude   -2.176e-01  3.422e-02 -6.358 2.32e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4356 on 3297 degrees of freedom
## (139 observations deleted due to missingness)
## Multiple R-squared:  0.08547, Adjusted R-squared:  0.08325
## F-statistic: 38.52 on 8 and 3297 DF, p-value: < 2.2e-16
```

```
educ.hyd50.cl=cluster.vcov(educ.hyd50, rd50.hb$dist_name)
educ.hyd50.se=sqrt(diag(educ.hyd50.cl)) #cluster standard errors
```

```
stargazer(educ.mys5, educ.mys10, educ.mys25, educ.mys50,
educ.hyd5, educ.hyd10, educ.hyd25, educ.hyd50,
se=list(educ.mys5.se, educ.mys10.se, educ.mys25.se, educ.mys50.se,
educ.hyd5.se, educ.hyd10.se, educ.hyd25.se, educ.hyd50.se), digits=3,
omit=c("TOT_POP", "TOT_SC", "TOT_ST", "Slope", "TerrainRug", "Latitude", "Longitude"),
dep.var.labels="Education: High Schools Availability",
column.labels =c("bw:10km", "bw:20km", "bw:50km", "bw:100km",
"bw:10km", "bw:20km", "bw:50km", "bw:100km"),
covariate.labels = c("Indirect Rule (Mysore)", "Indirect Rule (Hyderabad)", "Constant"),
add.lines = list(c("Controls", "\\checkmark", "\\checkmark", "\\checkmark", "\\checkmark",
"\\checkmark", "\\checkmark", "\\checkmark", "\\checkmark"),
omit.stat = c("rsq", "f", "adj.rsq", "ser"))
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Fri, Jun 16, 2023 - 15:15:41
## \begin{table}[!htbp] \centering
## \caption{}
## \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccccccc}
## \hline \hline \hline \hline
## & \multicolumn{8}{c}{\textit{Dependent variable:}} \\\
## \cline{2-9}
## \hline & \multicolumn{8}{c}{Education: High Schools Availability} \\\
## & bw:10km & bw:20km & bw:50km & bw:100km & bw:10km & bw:20km & bw:50km & bw:100km \\\
## \hline & (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8) \\\
## \hline
## Indirect Rule (Mysore) & 0.036$^{***}$ & 0.006 & $-$0.007 & 0.009 & & & & \\\
## & (0.015) & (0.031) & (0.042) & (0.038) & & & & \\\
```

```

## & & & & & & & \\\
## Indirect Rule (Hyderabad) & & & & & 0.085$^{*}$ & 0.087$^{*}$ & 0.112$^{*}$ & 0.128$^{***}$ \\\
## & & & & (0.048) & (0.052) & (0.057) & (0.044) \\\
## & & & & & & & \\\
## Constant & $-$6.262 & $-$7.310$^{**}$ & $-$8.219$^{**}$ & $-$3.100 & 29.315$^{***}$ & 20.684$^{*}$ \\\
## & (7.282) & (3.128) & (3.195) & (2.267) & (11.174) & (11.704) & (9.891) & (5.215) \\\
## & & & & & & & \\\
## \hline \\\[-1.8ex]
## Controls & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\\
## Observations & 595 & 1,100 & 2,404 & 4,221 & 551 & 901 & 1,787 & 3,306 \\\
## \hline
## \hline \\\[-1.8ex]
## \textit{Note:} & \multicolumn{8}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\\
## \end{tabular}
## \end{table}

```